

УДК 004.81

КЛАССИФИКАЦИЯ ТЕКСТОВ В СИСТЕМЕ УЗЛОВ ЛЕКСИЧЕСКОЙ ОНТОЛОГИИ

С.И. Данченков, В.Н. Поляков

Аннотация

В статье рассматривается проблема семантического индексирования текстов в терминах задачи классификации текстов по узлам лексической онтологии. Предлагается использовать данный подход в задачах идентификации объектов в полнотекстовых документах для повышения качества поиска информации. Приведены основные отличия данного подхода от задач разрешения многозначности и индексирования. Предложен метод классификации OntoKlass. Сделана формальная постановка задачи. Выполнена инженерная реализация метода.

Ключевые слова: искусственный интеллект, компьютерная лингвистика, классификация текстов, лексическая онтология, интеллектуальный анализ текстов.

Введение

Автоматическая классификация текстов представляет собой один из наиболее хорошо изученных разделов компьютерной лингвистики. Существуют солидные обзоры, дающие представления о результатах, полученных в этой области за последние годы [1]. Хорошо известны классические коллекции текстов (Reuters-21578 и др.), которые стали стандартом де-факто в области оценки качества методов и приложений для автоматической классификации текстов. Под автоматической классификацией (рубрикацией) информации понимается отнесение порции информации к одной или нескольким категориям из конечного множества рубрик [1, 2]. Она может применяться в следующих сферах:

- системы документооборота для классификации технической документации;
- библиотечные системы: классификация библиотечных источников;
- интернет-каталоги;
- поисковые технологии;
- фильтрация спама;
- фильтрация/отбор экономических новостей;
- структурирование новостей.

Существует ряд методов классификации больших объемов текстовой информации, в основе которых лежат технологии компьютерной лингвистики и алгоритмов распознавания образов.

Среди задач классификации выделяют:

- тематическую классификацию – задача: определить тему текста;
- жанровую классификацию – задача: определить жанр текста;
- региональную – задача: определить регион к которому относится текст;
- по источнику информации – задача: определить источник информации;
- по адресату информации – задача: определить адресат информации;
- по сектору экономики – задача: определить к какому сектору экономики относится информация.

Список может быть продолжен в зависимости от поставленных задач.

Кроме того, классификация бывает точной, когда каждой паре документ-категория сопоставляется булево значение 0 или 1 (соответствует документ категории или нет), и ранжированной (определяется степень принадлежности документа к категории). Известно несколько методов построения и обучения классификатора: Rocchio, разрешающие деревья, метод k соседей, метод опорных векторов (SVM) и др. [1]. Показатели (качество, цена, скорость) автоматической классификации постоянно улучшаются. Это обусловлено как применением новых методов и алгоритмов или улучшением уже имеющихся (например, добавление в модель эвристик [3]), так и развитием компьютерных технологий, следовательно, и вычислительных мощностей. Процесс автоматической классификации не предусматривает помощь специально обученных экспертов¹, что существенно снижает затраты и снижает риски по человеческому фактору. Важным моментом является наличие признанных метрик, позволяющих объективно оценить качество классификации [4].

Наряду с классификацией существуют и другие методы повышения качества поиска информации. В настоящее время перспективы организации более качественного, содержательного информационного поиска в сети Интернет связываются с разработкой онтологий [5, 6].

На формальном уровне онтология – это система, состоящая из набора понятий и набора утверждений об этих понятиях, на основе которых можно строить классы, объекты, отношения, функции и теории [5, 7, 8]. Стоит отметить, что существуют различные интерпретации определения онтологии среди специалистов по искусственному интеллекту [9–11] и в комьюнити по компьютерной лингвистике [12]. Основными компонентами онтологии являются классы или понятия, отношения, функции, аксиомы, примеры.

В зависимости от специфики декомпозиции предметной области существует несколько типов онтологий: верхнего уровня, предметных областей, прикладные онтологии, лексические онтологии [12]. В предложенном далее подходе мы будем использовать лексические онтологии.

Лексические онтологии – это лексико-семантические структуры, в которых установлены семантические связи между словами и словосочетаниями. При этом отличительной чертой онтологий является то, что узлами этой сети являются не сами слова (лексемы), а их семантические значения [7], то есть лексическая онтология представляет собой сеть вербализованных концептов (концепты с именами).

В качестве примера лексической онтологии можно привести ресурс WordNet [13]. Этот ресурс представляет в виде иерархической структуры систему значений словаря общеупотребительных слов английского языка. Вместе с тем возникает достаточно много проектов, которые описывают на основе модели WordNet терминологические системы конкретных предметных областей, то есть создают лексические онтологии этих областей [5, 11, 14, 15].

В настоящей работе рассматривается возможность применения технологий классификации и онтологических систем в рамках задачи классификации текстов с привязкой к узлам лексической онтологии. В этом случае каждый узел онтологии можно рассматривать как рубрику, к которой можно привязать текст, то есть фактически классифицировать его по узлам онтологий. В работе рассматривается прикладная онтология, включающая не только абстрактные, но и индивидуальные объекты (например, конкретные промышленные предприятия, персоналии топ-менеджеров). При этом решается относительно частная задача классификации текстов по этим терминальным узлам онтологии.

¹ В технологии машинного обучения труд экспертов используется при разметке обучающей выборки.

Отправной точкой в постановке задачи классификации текстов к узлам лексической онтологии является одна из задач информационного поиска и извлечения информации из текстов на естественном языке – задача выявления индивидуальных объектов в тексте. Под индивидуальным объектом будем понимать концепт, характеризующийся именем собственным, причем в тексте индивидуальный объект может иметь различные имена собственные, что несколько затрудняет его выявление, но дополнительная информация способна уточнить, о каком именно объекте идет речь (Пример 1).

Пример 1

*Открытое акционерное общество «Магнитогорский металлургический комбинат» [w₁₁] является крупнейшим предприятием чёрной металлургии [w₃₁] России, его доля в объёме металлопродукции [w₃₂], реализуемой на внутреннем рынке страны, составляет около 20%. ОАО «ММК» [w₁₂] входит в число двадцати крупнейших сталелитейных компаний [w₃₃] мира (по данным журнала *Metal Bulletin* [w₂₁]). Магнитогорский комбинат [w₁₃] представляет собой крупный металлургический комплекс [w₃₄] с полным производственным циклом, начиная с подготовки железорудного сырья [w₃₅] и заканчивая глубокой переработкой черных металлов [w₃₆]. По итогам 2005 года выручка от продажи товаров, продукции, работ, услуг ММК [w₁₄] составила 147 329.4 млн. рублей, прибыль до налогообложения – 39 744.7 млн. рублей.*

В данном примере:

- w₁₁, w₁₂, w₁₃, w₁₄ принадлежат микрословарю идентификаторов объекта;
- w₂₁ принадлежит микрословарю тематической лексики;
- w₃₁, w₃₂, w₃₃, w₃₄ принадлежат микрословарю индивидуальных лексических единиц².

Кроме того, индивидуальный объект может не всегда быть упомянут в явном виде (Пример 2). Однако в тексте может присутствовать информация, неявно указывающая на объект. Например, если речь идет о компаниях, упомянутые имена топ-менеджеров, названия партнеров или даже географические названия могут указывать на то, что в данном тексте речь идет именно об этой компании, а не о какой-то другой.

Пример 2

По решению губернатора Петра Сумина [w₂₁] Виктор Рашников [w₂₂] удостоен высшей награды Челябинской области [w₂₃] за выдающиеся заслуги в развитии металлургической промышленности [w₃₁] в регионе. Так оценен его труд в качестве руководителя крупнейшего в регионе и стране металлургического предприятия [w₃₂].

Здесь:

- w₂₁, w₂₂, w₂₃ принадлежат микрословарю тематической лексики;
- w₃₁, w₃₂ принадлежат микрословарю индивидуальных лексических единиц.

Необходимо подчеркнуть, что в нашем случае задача классификации текстов по узлам онтологической системы не сводится к задаче разрешения лексической многозначности, хотя иногда это можно сделать. В общем случае термин, которым обозначен индивидуальный объект в лексической онтологии, в тексте может и не

²Подробнее это рассматривается в разд. 1.

встретиться. Таким образом, текст может быть на основании косвенных признаков привязан к данному узлу онтологической сети. В этом и заключается новизна подхода.

1. Постановка задачи

1.1. Формальное описание онтологии. Будем представлять онтологию как пятерку вида:

$$O = (K, L, R_{kl}, R_{kk}, R_{ll}), \quad (1)$$

где K – множество концептов³, L – множество лексических единиц⁴, R_{kl} – множество отношений «концепт – лексическая единица»⁵, R_{kk} – множество семантических связей⁶ между концептами, R_{ll} – множество лингвистически обусловленных связей между лексическими единицами⁷.

1.2. Формальная постановка задачи. Формальная постановка задачи классификации в соответствии с [1] выглядит следующим образом. Имеется множество K узлов онтологической системы, множество D документов, необходимо построить отображение $C(K, D) \subset K \times D$, удовлетворяющее условию P .

1.3. Описание метода OntoKlass. Метод OntoKlass базируется на технологии машинного обучения и использует контролируемые⁸ словари лексических единиц четырех типов:

1) микрословарь идентификаторов объекта

$$V_1 = \{w_{11}, w_{12}, \dots, w_{1n}\}; \quad (2)$$

2) микрословарь тематической лексики

$$V_2 = \{w_{21}, w_{22}, \dots, w_{2m}\}; \quad (3)$$

3) микрословарь индивидуальных лексических единиц (имена сотрудников, географические названия)

$$V_3 = \{w_{31}, w_{32}, \dots, w_{3p}\}; \quad (4)$$

4) микрословарь уникальных лексических единиц (логотипы и названия продуктов)

$$V_4 = \{w_{41}, w_{42}, \dots, w_{4q}\}. \quad (5)$$

Микрословари V_1, V_2, V_3, V_4 формируются в процессе машинного обучения.

³В сети WordNet концепт описывается набором синонимов (синсетом). Пример: {завод, фабрика, предприятие}

⁴Лексическими единицами являются слова, устойчивые словосочетания или другие единицы языка, способные обозначать предметы, явления, их признаки и т. п.

⁵Это отношение описывает значение лексической единицы. Например: сосуд – {капилляр}, сосуд – {кувшин}.

⁶Примеры связей между концептами: связь «часть – целое»: отрасль – предприятие, предприятие – подразделение, предприятие – персонал; родо-видовая связь: ВУЗ – университет; синонимическая связь: помидор – томат; ассоциативная связь: студент – экзамен – профессор – аудитория.

⁷Примером могут служить морфологические формы: отпуск – отпускать.

⁸Термин обозначает словари, введенные вручную или сформированные автоматически в процессе машинного обучения.

Опишем текст как массив нормализованных словоформ⁹

$$D = \{w_{d1}, w_{d2}, \dots, w_{df}\}. \quad (6)$$

На этапе классификации формируются элементы A_1, A_2, A_3, A_4, A_5 алгоритма принятия решения:

$$A_1 = K_1 \cdot N_1 / N_{1\max}, \quad (7)$$

где $N_1 = |D \cap V_1|$, K_1 – весовой коэффициент, $N_{1\max} = |V_1|$;

$$A_2 = K_2 \cdot N_2 / N_{2\max}, \quad (8)$$

где $N_2 = |D \cap V_2|$, K_2 – весовой коэффициент, $N_{2\max} = |V_2|$;

$$A_3 = K_3 \cdot N_3 / N_{3\max}, \quad (9)$$

где $N_3 = |D \cap V_3|$, K_3 – весовой коэффициент, $N_{3\max} = |V_3|$;

$$A_4 = K_4 \cdot N_4 / N_{4\max}, \quad (10)$$

где $N_4 = |D \cap V_4|$, K_4 – весовой коэффициент, $N_{4\max} = |V_4|$;

$$A_5 = (A_1 + A_2 + A_3 + A_4) / 4. \quad (11)$$

В процессе классификации происходит наполнение матрицы отношений документ – рубрика:

$$C = \begin{array}{ccccc|c} & d_1 & d_2 & d_3 & \dots & d_y & \\ c_{11} & c_{12} & c_{13} & \dots & c_{1y} & k_1 \\ c_{21} & c_{22} & c_{23} & \dots & c_{2y} & k_2 \\ c_{31} & c_{32} & c_{33} & \dots & c_{3y} & k_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ c_{z1} & c_{z2} & c_{z3} & \dots & c_{zy} & k_z \end{array}, \quad (12)$$

где

$$C_{kd} \in \{0, 1\}. \quad (13)$$

Имеется система (массив) пороговых значений $T = \{T_1, T_2, T_3, T_4, T_5\}$. Таким образом, можно построить пространство правил для выполнения условий классификации, то есть привязки документа к узлу онтологии (см. табл. 1).

Правило принятия решения выбирается следующим образом:

$$C_{kd} = \begin{cases} 1, & \text{если } U \text{ выполняется;} \\ 0, & \text{если } U \text{ не выполняется.} \end{cases} \quad (14)$$

2. Инженерная реализация

Нами разработана программа, предназначенная для отнесения текстовых документов к искомому объекту, то есть она является по сути своей классификатором. Особенностью применяемого нами подхода является возможность привязки текста к объекту даже в тех случаях, когда объект упомянут в тексте иным термином или вовсе не упомянут в явном виде (только по косвенным признакам можно судить о принадлежности текста к объекту). В программе предусмотрены:

⁹Такое представление может быть получено в результате морфологического анализа текста в процессе лемматизации (получения первоначальной формы слова, исходя из других его словоформ).

Табл. 1

Правило	$A_1 > T_1$	$A_2 > T_2$	$A_3 > T_3$	$A_4 > T_4$	$A_5 > T_5$	Общая формула для U
Мягкое	Не важно	Не важно	Не важно	Не важно	И	$U: A_5 > T_5$
	И	Л	Л	Л	И	$U: (A_1 > T_1) \wedge (A_5 > T_5)$
	Л	И	Л	Л	И	$U: (A_2 > T_2) \wedge (A_5 > T_5)$
	Л	Л	И	Л	И	$U: (A_3 > T_3) \wedge (A_5 > T_5)$
	Л	Л	Л	И	И	$U: (A_4 > T_4) \wedge (A_5 > T_5)$
	И	И	Л	Л	И	$U: (A_1 > T_1) \wedge (A_2 > T_2) \wedge (A_5 > T_5)$
	И	Л	И	Л	И	$U: (A_1 > T_1) \wedge (A_3 > T_3) \wedge (A_5 > T_5)$
	И	Л	Л	И	И	$U: (A_1 > T_1) \wedge (A_4 > T_4) \wedge (A_5 > T_5)$
	Л	И	И	Л	И	$U: (A_2 > T_2) \wedge (A_3 > T_3) \wedge (A_5 > T_5)$
	Л	И	Л	И	И	$U: (A_2 > T_2) \wedge (A_4 > T_4) \wedge (A_5 > T_5)$
	Л	Л	И	И	И	$U: (A_3 > T_3) \wedge (A_4 > T_4) \wedge (A_5 > T_5)$
	И	И	И	Л	И	$U: (A_1 > T_1) \wedge (A_2 > T_2) \wedge (A_3 > T_3) \wedge (A_5 > T_5)$
	И	И	Л	И	И	$U: (A_1 > T_1) \wedge (A_2 > T_2) \wedge (A_4 > T_4) \wedge (A_5 > T_5)$
	И	Л	И	И	И	$U: (A_1 > T_1) \wedge (A_3 > T_3) \wedge (A_4 > T_4) \wedge (A_5 > T_5)$
	Л	И	И	И	И	$U: (A_2 > T_2) \wedge (A_3 > T_3) \wedge (A_4 > T_4) \wedge (A_5 > T_5)$
Жесткое	И	И	И	И	И	$U: (A_1 > T_1) \wedge (A_2 > T_2) \wedge (A_3 > T_3) \wedge (A_4 > T_4) \wedge (A_5 > T_5)$

где И – истина; Л – ложь.

- подключение коллекции текстовых документов с возможностью просмотреть список;
- ввод искомого запроса в поисковой строке, причем есть возможность сделать выбор из всех имеющихся в системе объектов;
- управление онтологией (редактирование узлов) с возможностью сохранять/загружать онтологию в файл/из файла;
- задание каждому объекту свойства типа индивид/множество;
- просмотр микрословарей выбранного объекта;
- настройка микрословарей объекта, реализована возможность сохранять/загружать микрословари в файл/из файла;
- извлечение терминов из обучающего образца (эталона);
- ручное добавление термина в микрословари;
- проверка на дублирование терминов в микрословарях;
- вывод результатов в виде списка, с возможностью просмотра текстов релевантных документов.

Главное меню программы (рис. 1) состоит из трех основных частей: «Верхнее меню» (меню «Файл» и «Помощь»), «Строка запроса» и «Результат выполнения запроса», в котором располагаются две области: список релевантных документов и текст выбранного документа. Пользователь вводит запрос (при этом система предлагает выбор из проиндексированных объектов), результатом выполнения которого будет список релевантных документов с возможностью просмотра текста

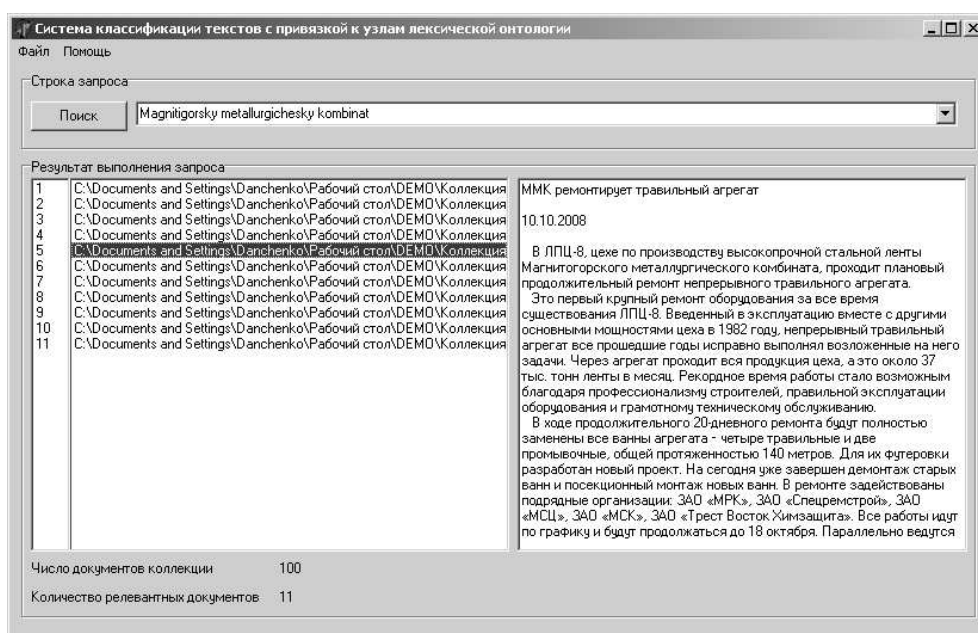


Рис. 1. Главное меню программы

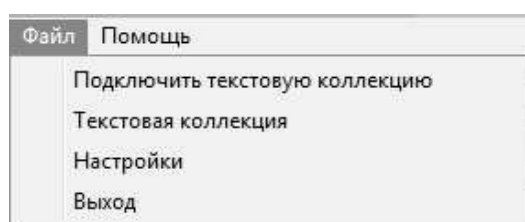


Рис. 2. Окно меню «Файл»

выбранного из данного списка документа. На экран выводится также информация о количестве документов в системе и числе документов, удовлетворяющих запросу.

В «Верхнем меню» имеется возможность подключить новую или просмотреть уже подключенную коллекцию текстовых документов. Для гибкого управления системой и оптимизации ее работы под нужды конкретного пользователя предусмотрено меню настроек (меню «Файл» – «Настройки» (рис. 2)).

Настройки распределены между двумя окнами: «Навигация по онтологии» (рис. 2) и «Настройка микрословарей объекта¹⁰» (рис. 3). В окне «Навигация по онтологии» пользователь может просмотреть ветвь онтологии с возможностью ее редактирования (добавлять/удалять узлы) и сохранения изменений. Кроме того, пользователь должен указать типы объектов системы (индивид/множество).

Микрословари выбранного объекта выводятся в отдельном окне в виде списков. Настройка микрословарей объекта происходит в одноименном окне (рис. 4). Загружается эталонный образец (текст по которому происходит обучение). Образец можно использовать как для наполнения всех четырех микрословарей (V_1 , V_2 , V_3 , V_4), так и лишь тех, что выбрал пользователь. Далее формируется список

¹⁰Под объектом понимается узел онтологии.

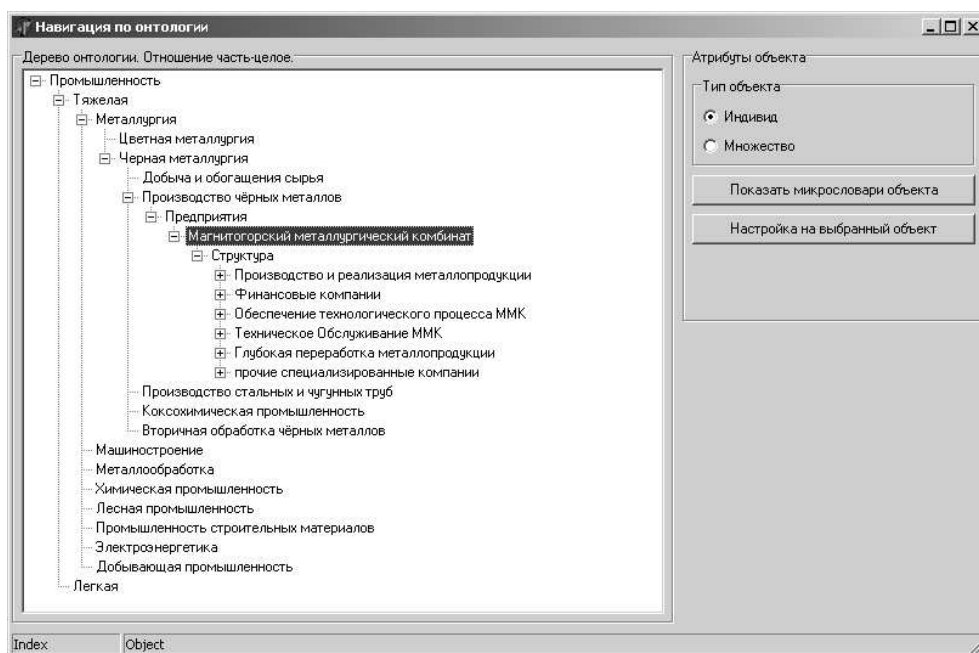


Рис. 3. Окно «Навигация по онтологии»

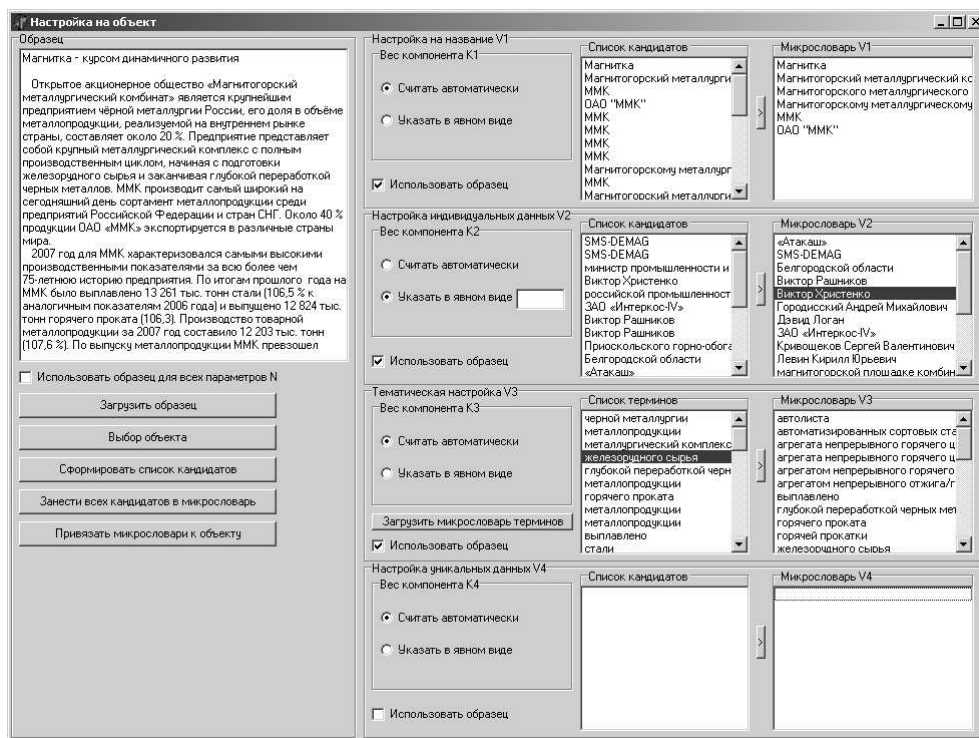


Рис. 4. Окно «Настройка микрословарей объекта»

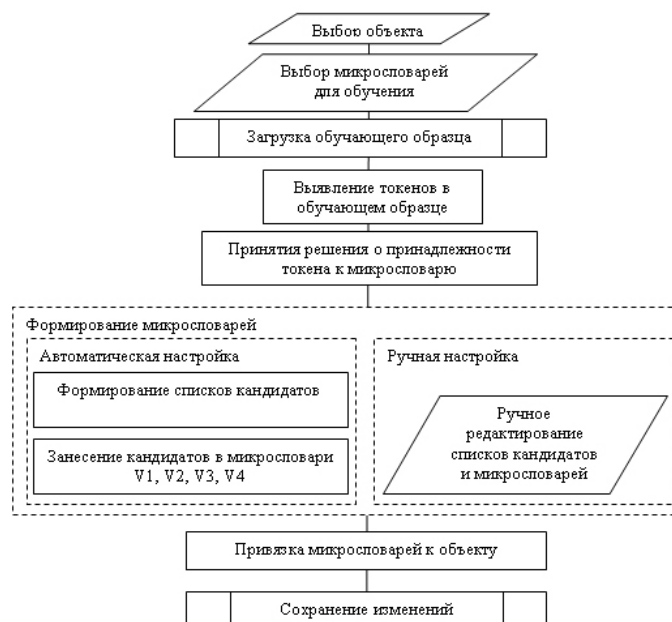


Рис. 5. Блок-схема обучения

кандидатов слов и словосочетаний для каждого микрословаря, после чего пользователь может занести в микрословарь либо всех кандидатов, либо кандидатов по выбору. При переносе кандидатов в микрословарь система автоматически исключает повторы. Предусмотрена возможность добавлять/удалять термин из списка кандидатов и непосредственно из микрословаря. Важной особенностью системы является гибкая настройка весов компонентов (K_1 , K_2 , K_3 , K_4). Для каждого компонента вес может быть определен автоматически или задаваться вручную.

Блок-схема обучения показана на рис. 5.

После формирования микрословарей они привязываются к объекту, на который происходила настройка. В окне навигации по онтологии у данного узла можно будет увидеть изменения в привязанных к нему микрословарях. Теперь пользователь может входить в меню поиска (Главное меню) и задавать запрос, который будет выполняться уже с учетом изменений.

Общая блок-схема, блок-схема расчета N компонентов и схема данных показаны на рис. 6–8).

Заключение

Результатом работы явилось разработка методики, формальной постановки задачи и инженерная реализация для задачи классификации текстов по узлам лексической онтологии. Данный метод может быть использован в целях повышения качества поиска информации, а также для частной задачи идентификации объекта.

Стоит выделить следующие основные отличия задачи классификации по узлам лексической онтологии от задачи разрешения многозначности и задачи индексирования.

1. Значение слова описано в словаре и, как правило, представляет собой концепт-класс, тогда как в нашем случае объектом может быть индивид.

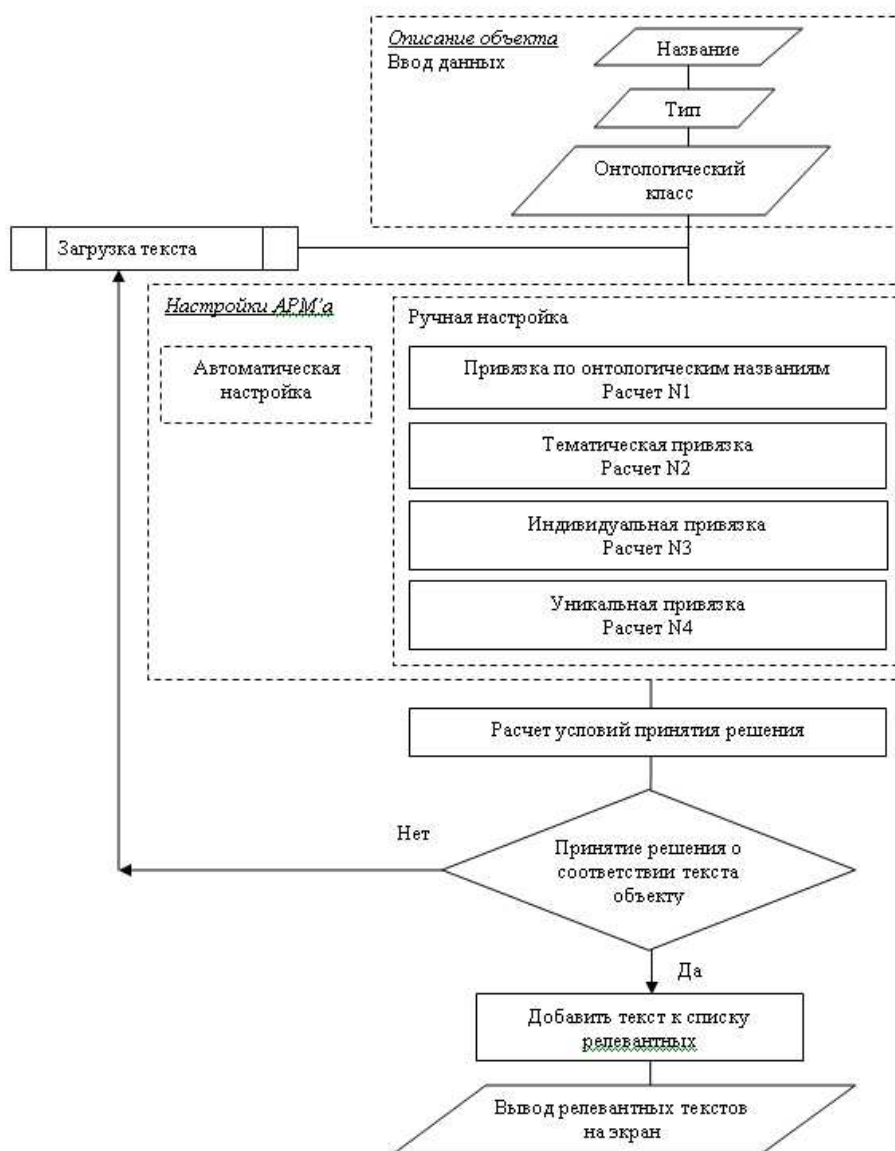


Рис. 6. Общая блок-схема

2. В нашей задаче классификации используются маркеры, связанные с жизненным циклом объекта, его функционирование, то есть это в какой-то степени максимально конкретизированный портрет предметной области, в которой функционирует объект, в то время как в задаче многозначности в основном используются доминирующие лексико-синтаксические связи. Таким образом, объект можно рассматривать как референт значения (концепта).

3. Задача классификации текстов с привязкой к узлам лексической онтологии отличается от задачи индексирования текстов по лексическим единицам тем, что в общем случае онтологический концепт (обозначающий класс или индивид) может быть в тексте не лексикализован в явном виде. Кроме того, задача индексирования обычно не решает проблем, связанных с омонимией и синонимией.

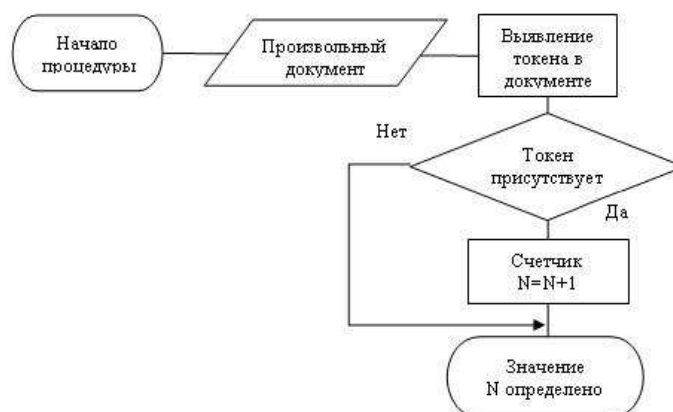


Рис. 7. Блок схема расчет N компонентов

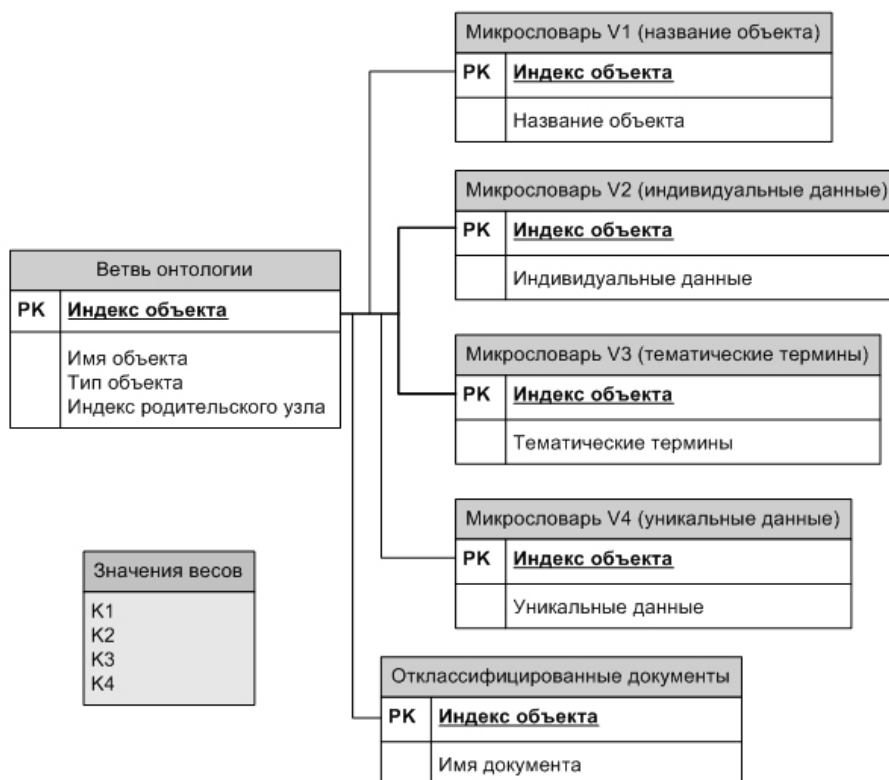


Рис. 8. Схема данных

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований и Правительства Республики Татарстан (проект № 09-07-97007-р-Поволжье_а).

Summary

S.I. Danchenkov, V.N. Polyakov. Automatic Text Classification in the System of Concepts Lexical Ontology.

In this work we consider a problem of text semantic indexing from positions of text classification problem with lexical ontology units. We offer to use this approach for objects identification problem and to improve quality of information retrieval. Basic differences of classification problem with using lexical ontology units from the text indexing task and from WSD task are formulated. A new method of classification is offered named as OntoKlass. Formal statement of the problem is made. Engineering realization of the method is executed.

Key words: artificial intelligence, computer linguistics, text classification, lexical ontology, text mining.

Литература

1. *Fabrizio S.* Machine Learning in Automated Text Categorization // ACM Computing Surveys. 2002. – V. 34, No 1. – P. 1–47.
2. *Агеев М.С., Добров Б.В., Лукашевич Н.В.* Автоматическая рубрикация текстов: методы и проблемы // Учен. зап. Казан. ун-та. Сер. Физ.-матем. науки. – 2008. – Т. 150, кн. 4. – С. 25–40.
3. *Polyakov V., Sidyakin O., Sinitcin V., Ten A.* Analysis of influence of heuristics on text classification effectiveness // Paper Collection “Text Processing and Cognitive Technologies” / Ed. by V. Solovyev, V. Polyakov, V. Goldberg. – Moscow-Kazan-Varna: Ucheba, 2005. – No 11. – P. 121–135.
4. *Данченков С.И., Поляков В.Н., Сидякин О.А.* Использование технологии классификации текстов Rubrix в задаче разрешения лексической многозначности // Труды Казанской школы по компьютерной и когнитивной лингвистике TEL-2006. – Казань, 2007. – С. 16–23.
5. *Добров Б.В., Лукашевич Н.В., Синицын М.Н., Шапкин В.Н.* Разработка лингвистической онтологии по естественным наукам для решения задач информационного поиска // Электронные библиотеки; перспективные методы и технологии, электронные коллекции: Труды седьмой Всерос. науч. конф. RCDL'2005 (Ярославль, 4–6 окт. 2005 г.). – Ярославль: Яросл. гос. ун-т, 2005. – С. 70–79.
6. *Россеева О.И., Загоруйко Ю.А.* Организация эффективного поиска на основе онтологий // Труды междунар. семинара Диалог'2001 по компьютерной лингвистике и ее приложениям. – Аксаково, 2001. – Т. 2. – С. 333–342.
7. *Поляков В.Н.* Использование технологий, ориентированных на лексическое значение, в задачах поиска и классификации // Проблемы прикладной лингвистики: Сб. ст. ИЯ РАН. – М.: Азбуковник, 2004. – Вып. 2. – С. 101–117.
8. *Гаврилова Т.А.* Онтологический подход к управлению знаниями при разработке корпоративных информационных систем // Новости искусственного интеллекта. – 2003. – № 2. – С. 24–30.
9. *Guarino N.* Formal Ontology and Information Systems // Guarino N. (ed.) Proc. of the 1st Intern. Conf. on Formal Ontologies in Information Systems, FOIS'98. – Trento, Italy: IOS Press, 1998. – P. 3–15.
10. *Hovy E.H.* Combining and standardizing large-scale, practical ontologies for machine translation and other uses // Proc. of the 1st Intern. Conf. on Language Resources and Evaluation (LREC). – Granada, Spain, 1998. – P. 535–542.
11. *Stumme G.* Using ontologies and formal concept analysis for organizing business knowledge // Wissensmanagement mit Referenzmodellen – Konzepte für die

- Anwendungssystem- und Organisationsgestaltung Heidelberg / Eds. J. Becker, R. Knackstedt. – Heidelberg: Physica, 2002. – P. 163–174.
12. *Соловьев В.Д., Добров Б.В., Иванов В.В., Лукашевич Н.В.* Онтологии и тезаурусы. – Казань, Москва: Казан. гос. ун-т, Моск. гос. ун-т, 2006. – 157 с.
 13. *Miller G., Beckwith R., Fellbaum C., Gross D., Miller K.* Introduction to WordNet: An On-Line Lexical Database // Intern. J. Lexicography. – 1990. – V. 3, No 4. – P. 235–312.
 14. *Roventini A., Marinelli R.* Extending the Italian WordNet with the Specialized Language of the Maritime Domain // Proc. of Second Intern. WordNet Conf. GWC. – 2004. – P. 193–198.
 15. *Vossen P.* Extending, Trimming and Fusing WordNet for Technical Documents // Proc. of WordNet and Other Lexical Resources: Applications, Extensions and Customizations. – Pittsburg, USA, 2001. – P. 125–131.

Поступила в редакцию
17.04.09

Данченков Сергей Игоревич – аспирант Московского института стали и сплавов.
E-mail: graysnpr@mail.ru

Поляков Владимир Николаевич – кандидат технических наук, доцент Московского государственного лингвистического университета и Московского института стали и сплавов, старший научный сотрудник Института языкознания РАН, г. Москва.
E-mail: pvn-65@mail.ru